

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 19, 2002, 12:44:21 : Search time 6219.51 Seconds
(without alignments)
2644.623 Million cell updates/sec

Title: US-10-003-759-3
Perfect score: 786
Sequence: 1 atgaacgtacgcgtcggt.....tttcgtaacggtgcagtag 786

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb_ba:*
- 2: gb_htg:*
- 3: gb_in:*
- 4: gb_em:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vl:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_om:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_pl:*
- 26: em_ro:*
- 27: em_sts:*
- 28: em_un:*
- 29: em_vl:*
- 30: em_htg_hum:*
- 31: em_htg_inv:*
- 32: em_htg_other:*
- 33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match Length	ID	Description

1	786	100.0	786	6	AX339682	AX339682 Sequence
2	786	100.0	1877	6	AX339680	AX339680 Sequence
3	746.8	95.0	1874	1	RMU72637	U72637 Rhodothermu
4	131.8	16.8	1116	1	AF233376	AF233376 Streptomy
5	131.8	16.8	1116	6	ARI29926	ARI29926 Sequence
6	131.8	16.8	1116	6	ARI68360	ARI68360 Sequence
7	105.8	13.5	1716	6	ARI68362	ARI68362 Sequence
8	97.8	12.4	3527	1	SHU51222	U51222 Streptomyce
9	96.2	12.2	1470	1	SREGLS	X7353 S.rochei eg
10	89	11.3	1311	1	AF130408	AF130408 Streptomy
11	87.6	11.1	1565	1	SLU04629	U04629 Streptomyce
12	87.6	11.1	41782	1	SGU11A	AL13210 Streptomy
13	69	8.8	11202	1	AE005123	AE005123 Halobacte
14	68.8	8.8	4262	1	AF335723	AF335723 Burkholde
15	64.4	8.2	14483	1	AE005992	AE005992 Mycobacte
16	64.4	8.2	67200	1	MTV017	AL021897 Mycobacte
17	57.6	7.3	1983	6	AX047403	AX047403 Sequence
18	55	7.0	36368	1	SC9B5	AL035206 Streptomy
19	54.8	7.0	346897	1	AP002995	AP002995 Mesorhizo
20	54.6	6.9	45313	1	SCD95A	AL357432 Streptomy
21	54.6	6.9	125020	9	AF429315	AF429315 Homo sapi
22	54.4	6.9	38995	1	SCF34	AL109974 Streptomy
23	54	6.9	34182	1	SC111	AL096849 Streptomy
24	53.6	6.8	77534	1	AF235504	AF235504 Streptomy
25	53	6.7	33779	1	SCGD3	AL096822 Streptomy
26	52.8	6.7	125020	9	AF429315	AF429315 Homo sapi
27	52.2	6.6	1849	1	SGSTRB1F	X78973 S.gallus DS
28	52	6.6	349116	1	AP003003	AP003003 Mesorhizo
29	51.4	6.5	3644	5	AF034576	AF034576 Gallus ga
30	51.2	6.5	22115	1	SCC22	AL096839 Streptomy
31	50.6	6.4	5525	1	SGSTRB1FG	X78974 S.-glaucesce
32	50.6	6.4	14596	1	AE007168	AE007168 Mycobacte
33	50.6	6.4	25459	1	SCAJ6985	XJ06985 Streptomy
34	50.6	6.4	30302	1	MTCV6G11	292774 Mycobacteri
35	50.6	6.4	40806	1	MSGY23	AD000016 Mycobacte
36	50	6.4	29825	1	SCD35	AL139178 Streptomy
37	50	6.4	36394	1	SCD35	AL160312 Streptomy
38	50	6.4	40104	1	SCF11	AL132662 Streptomy
39	49.8	6.3	36734	1	SCG10	AL049497 Streptomy
40	49.6	6.3	26555	1	SC2G1	AL391014 Streptomy
41	49.6	6.3	146555	2	OSJN00013	AL606448 Oryza sat
42	49.6	6.3	215050	1	AL646057	AL646057 Ralstonia
43	49.4	6.3	150814	2	AP003541	AP003541 Oryza sat
44	49.2	6.3	12130	1	AE005133	AE005133 Halobacte
45	49.2	6.3	41173	1	SCI41	AL132648 Streptomy

ALIGNMENTS

RESULT	1	AX339682	Sequence 3 from Patent WO0196382.	786 bp	DNA	linear	PAT 10-JAN-2002
LOCUS	AX339682	Sequence 3 from Patent WO0196382.					
DEFINITION	AX339682	Sequence 3 from Patent WO0196382.					
ACCESSION	AX339682	Sequence 3 from Patent WO0196382.					
VERSION	AX339682.1	GI:18135684					
KEYWORDS	AX339682.1	GI:18135684					
SOURCE	AX339682	Sequence 3 from Patent WO0196382.					
ORGANISM	AX339682	Sequence 3 from Patent WO0196382.					
REFERENCE	AX339682	Sequence 3 from Patent WO0196382.					
AUTHORS	AX339682	Sequence 3 from Patent WO0196382.					
TITLE	AX339682	Sequence 3 from Patent WO0196382.					
JOURNAL	AX339682	Sequence 3 from Patent WO0196382.					
FEATURES	AX339682	Sequence 3 from Patent WO0196382.					
BASE COUNT	AX339682	Sequence 3 from Patent WO0196382.					
ORIGIN	AX339682	Sequence 3 from Patent WO0196382.					

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 10:46:46 ; Search time 53.28 Seconds
(without alignments)
544.112 Million cell updates/sec

Title: US-10-003-759-2

Perfect score: 1446

Sequence: 1 MNVMAVLVLVLSLLFLGCDW.....PELMGGAGLSADFSVTVO 261

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A.Geneseq_032802.*

1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1318.5	91.2	260	AAV06369	Rhodothermus marinus
2	1318.5	91.2	260	AAV06369	Rhodothermus marinus
3	1318.5	91.2	260	AAV06369	Rhodothermus marinus
4	470.5	32.5	105	AAV06342	Amino acid sequenc
5	400.5	27.7	386	AAV067497	Rhodothermus marinus
6	396.5	27.4	371	AAV06367	Protien sequence o
7	396.5	27.4	371	AAV06367	Streptomyces sp. E
8	396.5	27.4	371	AAV06367	Actinomyces sp. 3
9	396.5	27.4	371	AAV06367	Streptomyces livid
10	396.5	27.4	371	AAV06367	Amino acid sequenc
11	343	23.7	381	AAV06368	Actinomyces cell
					Streptomyces livid

12	343	23.7	381	21	AAV06369	Rhodothermus marinus
13	330.5	22.9	429	21	AAV06369	Amino acid sequenc
14	255.5	17.7	261	20	AAV06369	Bacillus lichenifo
15	252.5	17.5	264	20	AAV06369	Erwinia carotovora
16	252.5	17.5	264	21	AAV06369	Erwinia carotovora
17	252.5	17.5	264	21	AAV06369	Amino acid sequenc
18	224.5	15.5	259	20	AAV06369	Aspergillus aculea
19	224.5	15.5	259	21	AAV06369	Aspergillus aculea
20	224.5	15.5	259	21	AAV06369	Amino acid sequenc
21	214	14.8	239	20	AAV06369	Aspergillus kawach
22	214	14.8	239	21	AAV06369	Aspergillus kawach
23	214	14.8	239	21	AAV06369	Amino acid sequenc
24	213.5	14.8	254	21	AAV06369	Humicola insolens
25	213.5	14.8	254	21	AAV06369	Humicola insolens
26	210.5	14.6	246	20	AAV06369	Actinomyces deserto
27	210.5	14.6	246	21	AAV06369	Actinomyces deserto
28	210.5	14.6	246	21	AAV06369	Amino acid sequenc
29	210.5	14.6	254	21	AAV06369	Humicola grisea EG
30	210.5	14.6	254	21	AAV06369	Humicola grisea EG
31	210.5	14.6	254	22	AAV06369	Humicola grisea en
32	210.5	14.6	254	22	AAV06369	Humicola grisea en
33	210	14.5	254	22	AAV06369	Humicola grisea en
34	203	14.0	255	20	AAV06369	Humicola insolens
35	200.5	13.9	348	20	AAV06369	Gliocladium roseum
36	200.5	13.9	348	21	AAV06369	Gliocladium roseum
37	200.5	13.9	348	21	AAV06369	Amino acid sequenc
38	199	13.8	104	20	AAV06369	Streptomyces sp. E
39	194	13.4	253	20	AAV06369	Humicola grisea EG
40	191.5	13.2	234	22	AAV06369	Trichoderma reesei
41	186.5	12.9	244	19	AAV06369	Tiarosporella phas
42	186.5	12.9	244	21	AAV06369	T. phaseolina xylo
43	186	12.9	244	20	AAV06369	Fusarium javanicum
44	186	12.9	244	21	AAV06369	Fusarium javanicum
45	186	12.9	244	21	AAV06369	Amino acid sequenc

ALIGNMENTS

RESULT: AAV06369
ID AAV06369 standard; Protein; 260 AA.

AC AAV06369;

DT 06-SEP-1999 (first entry)

XX Rhodothermus marinus EGIII-like cellulase.

DE Cellulase; endoglucanase; EGIII; textile; feed additive; baking; food processing; grain wet milling; pulp; paper.

XX Rhodothermus marinus.

XX Rhodothermus marinus.

XX Rhodothermus marinus.

PF 14-DEC-1998; 98MO-US26552.

PR 16-DEC-1997; 97US-0991720.

XX (CEMV) GENECOR INT INC.

PI Bower BS, Fowler T, Phillips JI;

XX WPI; 1999-395187/33.

DR EGIII like cellulase

PT Example; Fig 6; 47pp; English.

PS The present polypeptide represents a full-length sequence of a

CC

CC novel EGIII-like cellulase of Rhodothermus marinus. It was
 CC deduced from a gene sequence isolated from genomic DNA using PCR
 CC primers (see AY59180-91) based on conserved motifs (see AY06325-29)
 CC of Trichoderma reesei EGIII cellulase and related enzymes. PCR
 CC has been used to identify novel EGIII-like enzymes, including the
 CC present protein, from bacterial and fungal sources (see AY06331-70).
 CC The sequence shows homology to T. reesei EGIII (see AY06330). Also
 CC provided by the invention are vectors, host cells and methods
 CC for the recombinant production of such enzymes, which can be used
 CC in the treatment of cellulose-containing textiles, as feed
 CC additives, in the treatment of wood pulp, in the reduction of
 CC biomass to glucose, in the stone washing of indigo dyed denim, or
 CC as laundry detergent components (all claimed).
 XX
 XX Sequence 260 AA;

Query Match 91.2%; Score 1318.5; DB 20; Length 260;
 Best Local Similarity 90.3%; Pred. No. 1.7e-118;
 Matches 241; Conservative 4; Mismatches 9; Indels 13; Gaps 2;
 Qy 1 MMYMRAVLVLSLLFLGCDWLPFGDNGKEPEPEPTVELCGRWDAVDAGGRYRVINN 60
 Db 1 mmyravvlvlslllfgcdwlpfgdngkepepeptvelcgrwdardvaggryrvinn 60
 Qy 61 VMCAETACQIEVGLTGNFTITRADHNGNNVAAYPAIYFGCHWGA-----CTSNGLP 114
 Db 61 vmeatacqievgletgnftitradhngnnvaaypalyfychwaparairdcaaragav 120
 Qy 115 RRVQELSVRTSMWLTPTTGRNNAAYDIWFSPVNSGNGYSGGAELMWLNMGVMPG 174
 Db 121 rrahld-----vtpittgrwnaaydiwfspvntsgngysggaelmwlnwngvmpg 173
 Qy 175 GSRVATVELAGATWVYADWNNYIAYRRTTPTTSVELDLKAFIDDAVARGYRPEWY 234
 Db 174 gsrvatvelagatwvwyadwvnyayrrttpttsvelldkafiddavargyirpewy 233
 Qy 235 LHAVETGPELWEGGAGLSADFSVTQ 261
 Db 234 lhavetgfeiweggagirtadfsvtq 260

RESULT 2
 AAB14882
 ID AAB14882 standard; Protein; 260 AA.
 XX
 AC AAB14882;
 XX
 DT 21-NOV-2000 (first entry)
 XX
 DE Emericella desertoru EGIII-like cellulase.
 XX
 KW Emericella desertoru; Trichoderma reesei; endoglucanase III; EGIII;
 KW cellulase; mutant; enzyme stability; textile treatment;
 KW wood pulp treatment; feed additive; detergent.
 XX
 OS Emericella desertoru.
 XX
 PN WO200037614-A2.
 XX
 PD 29-JUN-2000.
 XX
 PF 12-NOV-1999; 99WO-US26704.
 XX
 PR 18-DEC-1998; 98US-0216295.
 XX
 PA (GENV) GENENCOR INT INC.
 XX
 PI Mitchinson C, Wendt DJ;
 XX
 DR WPI; 2000-482483/42.
 XX
 PT Novel endoglucanase III or endoglucanase III-like cellulase useful for

PT treating textiles and wood pulp comprises a substitution or deletion at
 PT specified positions in the wild form of endoglucanase III -
 XX Example 1; Fig 3; 52pp; English.
 XX
 CC The present sequence is a cellulase related to endoglucanase III (EGIII)
 CC from Trichoderma reesei. EGIII-like genes were isolated from genomic DNA
 CC libraries constructed from various microorganisms by PCR. The isolated
 CC genes showed significant homology to EGIII from T. reesei. Certain
 CC substitution and deletion mutations have been incorporated into EGIII and
 CC EGIII-like cellulases to produce variant enzymes with improved stability,
 CC e.g. increased resistance to temperature stress. The mutants may be used
 CC in textile and wood pulp treatment, as a feed additive, and for reducing
 CC biomass to glucose. They are also useful for stonewashing or indigo dyed
 CC denim and as an agent in laundry and dish detergents.

XX Sequence 260 AA;
 SQ
 Query Match 91.2%; Score 1318.5; DB 21; Length 260;
 Best Local Similarity 90.3%; Pred. No. 1.7e-118;
 Matches 241; Conservative 4; Mismatches 9; Indels 13; Gaps 2;
 Qy 1 MMYMRAVLVLSLLFLGCDWLPFGDNGKEPEPEPTVELCGRWDAVDAGGRYRVINN 60
 Db 1 mmyravvlvlslllfgcdwlpfgdngkepepeptvelcgrwdardvaggryrvinn 60
 Qy 61 VMCAETACQIEVGLTGNFTITRADHNGNNVAAYPAIYFGCHWGA-----CTSNGLP 114
 Db 61 vmeatacqievgletgnftitradhngnnvaaypalyfychwaparairdcaaragav 120
 Qy 115 RRVQELSVRTSMWLTPTTGRNNAAYDIWFSPVNSGNGYSGGAELMWLNMGVMPG 174
 Db 121 rrahld-----vtpittgrwnaaydiwfspvntsgngysggaelmwlnwngvmpg 173
 Qy 175 GSRVATVELAGATWVYADWNNYIAYRRTTPTTSVELDLKAFIDDAVARGYRPEWY 234
 Db 174 gsrvatvelagatwvwyadwvnyayrrttpttsvelldkafiddavargyirpewy 233
 Qy 235 LHAVETGPELWEGGAGLSADFSVTQ 261
 Db 234 lhavetgfeiweggagirtadfsvtq 260

RESULT 3
 AAY84347-
 ID AAY84347 standard; Protein; 260 AA.
 XX
 AC AAY84347;
 XX
 DT 12-JUL-2000 (first entry)
 XX
 DE Amino acid sequence of an endoglucanase III (EGIII)-like cellulase.
 XX
 KW Endoglucanase III; EGIII; EGIII-like cellulase; surfactant stability;
 KW cellulase; textile processing; textile cleaning; stonewashing;
 KW indigo dyed denim; cellulose containing fabric; fabric smoothing;
 KW pill removal; fibril removal; cotton; cellulosic fibre; dyeing; detergent;
 KW animal feed; wood pulp; paper; grain; biomass reduction; glucose.
 XX
 OS Rhodothermus marinus.
 XX
 PN WO2000014208-A1.
 XX
 PD 16-MAR-2000.
 XX
 PF 24-AUG-1999; 99WO-US19154.
 XX
 PR 03-SEP-1998; 98US-0146729.
 XX
 PA (GENV) GENENCOR INT INC.
 XX
 PI Fowler T;

L4 ANSWER 7 OF 10 CAPLUS COPYRIGHT 2003 ACS on STN DUPLICATE 1

ACCESSION NUMBER: 2001:454833 CAPLUS

DOCUMENT NUMBER: 135:45237

TITLE: **Deletion** of a cytotoxic, N-terminal putative signal peptide results in a significant increase in production yields in *Escherichia coli* and improved specific activity of Cell12A from *Rhodothermus marinus*

AUTHOR(S): Wicher, K. B.; Abou-Hachem, M.; Halldorsdottir, S.; Thorbjarnadottir, S. H.; Eggertsson, G.; Hreggvidsson, G. O.; Karlsson, E. Nordberg; Holst, O.

CORPORATE SOURCE: Department of Biotechnology, Center for Chemistry and Chemical Engineering, Lund University, Lund, 221 00, Swed.

SOURCE: Applied Microbiology and Biotechnology (2001), 55(5), 578-584

CODEN: AMBIDG; ISSN: 0175-7598

PUBLISHER: Springer-Verlag

DOCUMENT TYPE: Journal

LANGUAGE: English

AB The thermostable **cellulase** Cell12A from *Rhodothermus marinus* was produced at extremely low levels when expressed in *Escherichia coli* and was cytotoxic to the cells. In addn., severe aggregation occurred when moderately high concns. of the enzyme were heat-treated at 65.degree.C, the growth optimum of *R. marinus*. Sequence anal. revealed that the catalytic module of this enzyme is preceded by a typical linker sequence and a highly hydrophobic putative signal peptide. Two **deletion** mutants lacking this hydrophobic region were cloned and successfully expressed in *E. coli*. These results indicated that the N-terminal putative signal peptide was responsible for the toxicity of the full-length enzyme in the host organism. This was further corroborated by cloning and expressing the **hydrophobic N-terminal** domain in *E. coli*, which resulted in extensive cell lysis. The **deletion** mutants, made up of either the catalytic module of Cell12A or the catalytic module and the putative linker sequence, were characterized and their properties compared to those of the full-length enzyme. The specific activity of the mutants was approx. three-fold higher than that of the full-length enzyme. Both mutant proteins were highly thermostable, with half-lives exceeding 2 h at 90.degree.C and unfolding temps. up to 103.degree.C.